1/5/1
DIALOG(R)File 351:Derwent WPI

012567419 **Image available**
WPI Acc No: 1999-373526/199932
XRPX Acc No: N99-278866
Compression of documents with markup language preserving syntactical structure
Patent Assignee: MARTIN B K (MART-I); UNWIRED PLANET INC (UNWI-N)
Number of Countries: 028 Number of Patents: 004
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week	
EP 928070	A2	19990707	EP 98310574	A	19981222	199932	B
CN 1222009	A	19990707	CN 98119772	A	19980928	199945	
JP 11284517	A	19991015	JP 98367699	A	19981224	200001	
KR 99066882	A	19990816	KR 9860065	A	19981229	200045	

Priority Applications (No Type Date): US 97999518 A 19971229
Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
EP 928070	A2	E	20	H03M-007/30	
Designated States (Regional): AL AT BE CH CY DE DK ES FI FR GB GR IE IT LI LT LU LV MC MK NL PT RO SE SI					
CN 1222009	A			H03M-007/30	
JP 11284517	A		14	H03M-007/30	
KR 99066882	A			G06F-017/20	

Abstract (Basic): EP 928070 A2
NOVELTY - A client uses a network to access resources provided by servers (112), while a remote device (106) provides a user interface and a computer (110) exchanges information with the network. The remote device contains a display, buttons, storage, a transmitter and a receiver and may be a wireless telephone, while communication with the computer uses technology such as electromagnetic transmission in the radio frequency to infrared portions of the spectrum. A document is compressed into codes including an indication of the markup language tag and the content
DETAILED DESCRIPTION - An independent claim is included for a method of receiving decoded information representing a document
USE - Compression of information having syntactical structure such as document description representing a document
ADVANTAGE - Reduced bandwidth and resource requirements to compress expressed information
DESCRIPTION OF DRAWING(S) - The drawing is a schematic illustration of major components of system in which various aspects of present invention may be carried out
Servers (112)
Remote device (106)
Computer (110)
pp; 20 DwgNo 1/8
Title Terms: COMPRESS; DOCUMENT; LANGUAGE; PRESERVE; STRUCTURE
Derwent Class: T01; U21
International Patent Class (Main): G06F-017/20; H03M-007/30
International Patent Class (Additional): G06F-017/21; G06F-017/30
File Segment: EPI

[19]中华人民共和国国家知识产权局

[51]Int.Cl[6]
H03M 7/30

[12]发明专利申请公开说明书

[21]申请号 98119772.8

[43]公开日 1999年7月7日

[11]公开号 CN 1222009A

[22]申请日 98.9.28 [21]申请号 98119772.8
[30]优先权
[32]97.12.29 [33]US [31]999518
[71]申请人 小布鲁斯·K·马丁
地址 美国加利福尼亚州
[72]发明人 小布鲁斯·K·马丁

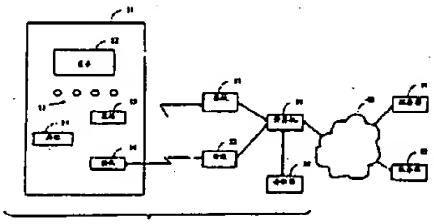
[74]专利代理机构 柳沈知识产权律师事务所
代理人 马 莹

权利要求书3页 说明书14页 附图页数6页

[54]发明名称 保留语法结构的利用标记语言压缩文档的方法

[57]摘要

一种标记语言文档压缩方法，用于在移动电话和英特网之间带宽有很的通信频 道上传送用标记语言表达的超文本文档。通过对文档元素进行压缩编码，使元 素的语法特征容易由编码表达式得出，并在与压缩成代码的元素的类型无关的 代码开头位置上传递标记语言标签属性和内容的语法信息存在的指示，可以有 效地处理文档元素压缩或编码表达式而不需要展开或解码。此外，利用现有编 码器和解码器可以有效处理它们不能识别的标记语言扩充部分。



ISSN1008-4274

权　利　要　求　书

1. 一种减少代表文档的输入信息的容量要求的方法，包括：

接收代表文档的输入信息并从中识别多个元素，其中每个元素有一个相应的类型，而且至少某些元素具有表示一个或更多相应语法特征的语法信息；

生成多个代码，各个代码有一个开头且表示各个元素的至少一部分，且所要求的信息容量低于所表示部分要求的信息容量，其中各个代码传递了各个元素类型及指示各个元素语法信息是否存在的语法指示，并且各个代码在与各代码开头相应的预定位置上传递该语法指示；及

通过将多个代码和不被多个代码表示的部分元素汇编成适于传输或贮存的形式从而生成表示文档的编码信息。

2. 根据权利要求 1 中所述的方法，其中，所述元素符合基于标签的标记语言，每个元素包含一个标记语言标签，所述语法信息包含标签属性和标签内容。

3. 根据权利要求 2 中所述的方法，其中，所述基于标签的标记语言符合标准通用标记语言(SGML)文档类型定义(DTD)。

4. 根据权利要求 2 中所述的方法，其中，生成的所述代码具有使所述语法指示以不依赖元素类型的方式指示语法信息是否存在的形式。

5. 根据权利要求 2 中所述的方法，其中，所述代码具有固定长度且在与所述代码开头相应的一个固定位置上传递语法指示。

6. 根据权利要求 1 中所述的方法，其中，所述代码具有固定长度且在与所述代码开头相应的一个固定位置上传递语法指示。

7. 根据权利要求 1 中所述的方法，还进一步包括：从许多代码簿中选择一本代码簿，其中至少某些代码根据所选中的代码簿而生成且所选中的代码簿的指示汇编在所述编码信息中。

8. 一种从编码信息中复原表示文档的解码信息的方法，其中，文档包含许多元素，此方法包括：

接收表示文档的编码信息并从中识别多个代码，其中各个代码有一个开头，代表各个元素的至少某个部分，传递了指示各个元素类型的各个类型指示并传递了指示语法信息是否存在的各个语法指示，所述语法信息表示所述

各个元素一个或更多的语法特征；

从各个代码的与各代码开头相应的预定位置上得到各个语法指示；

生成多个解码表达式，其中各个解码表达式从各自代码中得出和对应于由相关代码表示的相关元素的对应部分，其中各个语法指示控制表示语法信息的解码表达式的生成，并且所得到的各个解码表达式要求的信息容量大于相关代码要求的信息容量；及

将多个解码表达式和不用所述代码表示的部分元素进行汇编从而生成表示文档的输出信息。

9. 根据权利要求 8 中所述的方法，其中，所述元素符合基于标签的标记语言，每个元素包含一个标记语言标签，所述语法信息包括标签属性和标签内容。

10. 根据权利要求 9 中所述的方法，其中，所述基于标签的标记语言符合标准通用标记语言(SGML)文档类型定义(DTD)。

11. 根据权利要求 9 中所述的方法，其中，所述代码具有使所述语法指示以不依赖元素类型的方式指示语法信息是否存在的形式。

12. 根据权利要求 9 中所述的方法，还进一步包括：通过根据所述输出信息中的元素来处理输出信息，产生用于在一显示装置上显示的信号，其中，该处理过程使用所述输出信息中一个或更多元素的语法指示，来避免会影响一个或更多所述显示特征的语法信息处理。

13. 根据权利要求 9 中所述的方法，还进一步包括：通过根据所述编码信息中的代码来处理编码信息，产生用于在一显示装置上显示的信号，其中，该处理过程使用所述编码信息中一个或更多代码的语法指示，来避免会影响一个或更多所述显示特征的语法信息处理。

14. 根据权利要求 9 中所述的方法，其中，所述编码信息包含一个或更多无支持的代码样例，从这些代码中不能得到相应的解码表达式，并且所述输出信息也通过汇编一个或更多无支持的代码样例来生成。

15. 根据权利要求 9 中所述的方法，其中，所述代码具有固定长度且在与所述代码开头相关的固定位置上传递所述语法指示。

16. 根据权利要求 8 中所述的方法，其中，所述代码具有固定长度且在与所述代码开头相关的固定位置上传递所述语法指示。

17. 根据权利要求 8 中所述的方法，其中，所述编码信息包含了从许多

代码簿中选中的一个代码簿的指示，并且至少某些解码表达式是根据所选中的代码簿从所述代码导出的。

18.一种从包含许多压缩编码元素的编码信息中复原表示文档的解码信息的方法，包括：

处理编码元素从而识别元素类型并得到元素语法特征的语法指示，其中该语法指示从编码元素内与编码元素开头相应的预定位置上得出而且元素类型的压缩表达式展开为标记语言标签的解压缩形式；

如果该语法指示指示至少存在一个标签属性，则通过把标签属性信息的压缩表达式展开为标记语言标签属性名称或标签属性值的解压缩形式，来处理编码元素中的标签属性信息；和

如果语法指示指示存在标签内容，则根据适用于标签内容的处理过程处理编码元素中的标签内容信息。

19. 根据权利要求18中所述的方法，其中，所述标记语言标签符合标准通用标记语言(SGML)文档类型定义(DTD)。

20. 根据权利要求18中所述的方法，其中，所述编码元素具有使所述语法指示以不依赖于元素类型的方式指示语法信息是否存在的形式。

21. 根据权利要求18中所述的方法，还进一步包括：根据所述输出信息中的元素来处理输出信息，产生用于在一显示装置上显示的信号，其中该处理过程使用所述输出信息中的一个或更多元素的语法指示，来避免会影响一个或更多所述显示特征的标签属性信息或标签内容处理。

22. 根据权利要求18中所述的方法，其中，所述编码信息包含一个或更多无类型支持的编码元素样例，这种编码元素不能展开为标记语言标签的解压缩形式，并且所述输出信息也通过汇编一个或更多无类型支持的编码元素样例来生成。

23. 根据权利要求18中所述的方法，其中，所述编码信息包含从许多代码簿中选中的某个代码簿的指示，并且标记语言标签、标签属性名称或标签属性内容根据所选中的代码簿展开成解压缩形式。

24. 根据权利要求18中所述的方法，其中，所述元素类型的压缩表达式具有固定长度，并且在所述元素类型压缩表达式内某一固定位置上传递所述语法指示。

说　明　书

保留语法结构的
利用标记语言压缩文档的方法

本发明一般涉及用于在低带宽通信频道上向接收装置传输信息的信息压缩方法，具体涉及用于向类似手持移动电话的无线装置传输具有某种语法结构的信息（诸如符合通用标记语言规则的文档描述）的信息压缩方法。

象 Internet （英特网）这样的网络已经存在好多年了；然而，直到最近它们才成为信息交换的流行媒介，近来 Internet 的应用迅猛增长，从很大程度上缘于设备和方法的发展，简化了用户访问和阅读存贮于网络服务器的多媒体信息所需的操作过程。访问资源，如人们熟知的超链接(hyperlink)，允许不同的信息片不按顺序组织起来并使用户很方便地在链接的信息中浏览。通过给网络中可提供的不同的每一片多媒体信息分配唯一标识符，即大家所熟知的统一资源定位器(URL)，用户可以很方便地访问信息而不需考虑信息存贮的位置。参与这种“超媒体(hypermedia)”网络的网络客户机和服务器在此处分别称之为超媒体客户机和超媒体服务器。

促成这种增长的一个重要发展是诸如“标记语言(markup language)”和相关处理工具的使用，这些工具定义并采用了规定文档各种语法特征的多个元素。现在使用的许多标记语言符合国际标准 ISO 8879 : 1986，此标准定义了一系列在此称作“标准通用标记语言(SGML)”的基于标签语言(tag-based language)的基本规则。在 Internet 上使用最广泛的符合 SGML 规则的标记语言也许是超文本标记语言(HTML)。

用基于标签的标记语言表示的文档常常用称为浏览器或阅读器的应用软件来显示和操作。这些应用软件执行符合相应标记语言规则的处理，对表示文档的信息进行(语法)分析和解释，从而正确地显示文档内容。

符合 SGML 类标记语言的表示文档的信息，一般包含有标签和可能有的相关标签属性和标签内容的一些元素，这些元素传递了文档中携带信息的语法特征。

标签标识了元素类别，例如，在 HTML 中，表示整个文档的元素通过标注在文档开始和结束处的标签来标识，表示文本段落的元素通过标注在段

落起始处的标签来标识，将要被显示带有下划线的文本通过标注在下划线开始和结束的标签来标识。

标签属性提供了指定元素一个或多个特征的信息。例如，表示嵌入在文档中的图像文件的一个标签，包含了指定嵌入图像文件名称的属性。根据标记语言的规范，标签属性依据相应的标签类型可以是任选项或必选项。

标签内容一般表示用来显示或被用户可操作的信息。标签内容根据标签类别可以是任选或必选项，也可以包含其它本身也有标签、属性和内容的"嵌套"元素。

由于 SGML 本身有很大的灵活度，所以那些符合 SGML 的标记语言能为执行文档元素提供非常灵活和强大的工具。这种灵活度是有代价的。它需要额外的带宽来传递标签和标签属性及需要额外的资源对标签和标签属性进行(语法)分析和解释。在 HTML 中，标签和属性用字符串以一种类似于<tagid name = value>的形式表达，其中 tagid 是标签的标识符， name 是属性的名称， value 是分配给那个属性的值。一个标签可以有不只一个属性。

由于可以很容易得到具有充足计算能力和足够宽通信频道的个人计算机和工作站，因此在多数情形中，传递和处理标签和标签属性所需的额外带宽和资源并不是一个显著的不利因素。

然而，人们对通过移动装置，尤其是无线电话这类手持装置，来与 Internet 网这样的网络相连接的超媒体服务器进行访问的兴趣越来越大。这些装置在处理能力和存贮空间上都有严格的限制。此外，连接移动装置与其余网络的通信频道带宽也受到极严格的限制。

一部无线电话的资源只有一般台式或便携电脑所提供的资源中的一小部分。通常，处理能力不到多数计算机的百分之一，存贮空间一般远小于 150 千字节(kB)，通话路径通常在范围 400 ～ 19,200 比特/每秒，使用通信路径的费用以每 100kB 或更多 kB 多少美元计算。

通过减少通信频道上所传递信息的容量要求，可以有效地使用有限带宽的通信频道。采用一些数据格式或信息压缩可以减少信息的容量要求。

已经对一般目地的压缩方案如 Huffman 编码进行了研究，但不幸的是，由于一般目的的方案得到的压缩信息结果遮蔽了底层信息的语法特征而并不十分理想，换句话说，标签的识别及标签属性和内容的存在与否不容易从压缩表达式中求出，此外，一般目的压缩方案不能象基于某种特定的标记语

言的压缩方案那样减少很多信息的容量要求。

对基于类似 HTML 的特定标记语言的各种压缩方案进行了研究。此类压缩方案通过利用特定标记语言的已知特征能够达到更高的压缩程度。例如，某种标记语言的专门压缩方案不允许对无内容的标签传递标签内容的可能性。不幸的是，这些方案要求浏览器或展开过程能处理和扩展所有压缩元素。标记语言的扩充和改变部分不能从压缩表达式中复原，除非修整浏览器来处理新的语言特点；否则，新特征的压缩掩蔽了包含新特征的元素及嵌套元素的语法特征。特别是，既使在不能或无需使用新特征的某种应用程序或装置中加入浏览器，浏览器也必须被修整。

例如，假如某种基于标记语言的压缩方案被扩充来压缩某种新的显示格式，浏览器不能从压缩表达式中复原显示格式信息，除非将浏览器修整为包含扩展新特征所要求的处理程序。此外，没有这种修整，浏览器不能忽略或跳过新特征而去展开其余信息，因为它的处理能力不能够测定扩充的新压缩特征。

本发明目的是在不遮蔽底层文档元素的语法特征的情形下，减少传输和处理表达文档的信息所要求的带宽和资源。

根据本发明的一个特点，一种减少表达文档的输入信息所要求容量的方法，包括：接收输入信息且识别其中的多个元素，每个元素有相应的类型而且至少有些元素具有表示一个或更多相应语法特征的语法信息；生成多个代码，每个代码有开头部分并表示相应元素的至少一部分且代码要求的信息容量低于所代表部分要求的信息容量，每个代码传递各自元素类型及指示各自元素语法信息是否存在的语法指示，且每个代码在与各自代码开头部分相对应的预定位置传递语法指示；及通过将多个代码和多个元素中不用多个代码表示的部分汇编成适于传输和存贮的形式从而生成表示文档的编码信息。

根据本发明的另一特点，一种从编码信息中复原包含多个元素的文档的方法，包括：接收表示文档的编码信息并识别其中的多个代码，其中，每个代码有一个开头部分，表示相应元素的至少一部分，传递指示相应元素类型的类型指示和指示表示相关元素的一个或更多语法特征的语法信息是否存在的相关语法指示；从与各个代码开头部分相对应的预定位置得到各个语法指示；生成多个解码表达式，其中每个解码表达式由相应代码得出并对应于用每个代码表示的相应元素的部分，其中各语法指示控制表示语法信息的解

码表达式的生成且所得出的相应解码表达式的信息容量要求大于相应代码的信息容量要求；及将多个解码表达式与多个元素中不用代码表示的部分元素进行汇编从而生成表示文档的输出信息。

根据本发明的另一特点，一种从多个压缩的编码元素中复原文档的方法，包括：处理编码元素以识别元素类型并且得到元素语法特征的语法指示，其中语法指示由与编码元素开头相应的编码元素内预定位置得出，并且元素类型的压缩表示展开为标记语言标签的解压缩形式；如果语法指示指示至少一个标签属性存在，则通过将标签属性信息的压缩表达式展开成标记语言标签属性名称或标签属性值的解压缩形式，来处理编码元素中的标签属性信息；及如果语法指示指示标签内容存在，则根据适用于标签内容的处理过程处理编码元素的标签内容信息。

本发明及其优选实施例的各种特点可以通过附图及下面有关说明而得到更好地理解。其中在几幅附图中的相同标号表示相同单元。下面说明的内容及附图在此仅作为例子给出，而不该理解为对本发明范围的限制。

图 1 是某系统主要成份的示意图，其中本发明的各个特点可在该系统内体现。

图 2 是生成文档元素压缩表达式的处理或装置的方框图。

图 3 是从压缩表达式中复原文档元素的处理或装置的方框图。

图 4 是生成文档元素压缩表达式的处理的状态图。

图 5 是从压缩表达式中复原文档元素的处理的状态图。

图 6 是压缩或展开文档信息过程的功能流程图。

图 7 说明一个用标记语言表达的简单文档。

图 8 示出依据本发明的编码过程处理的表示图 7 中文档的编码信息。

综述

图 1 示出说明一个实现本发明各个特点的系统。图中画出的一些成份有可能在不同实施例中省略，如图所示，客户机 1 利用网络 40 访问由服务器 51 和服务器 52 提供的资源。尽管设想服务器 51 和服务器 52 是超媒体服务器，多半符合超文本传输协议(HTTP)，进行操作，但对实施本发明并不是必要的。在典型实施例中，远程终端 11 为用户提供了向用户发送信息和从用户接收输入的接口，计算机 31 以常规网络客户机的方式与网络 40 进行信息交换。

计算机 31 将参数与信息贮存在存贮器 32，存贮器 32 常常是随机存取存贮器(RAM)、只读存贮器(ROM)及磁盘和光盘驱动器这样的永久存贮装置的混合体。计算机 31 通过接收器 21 和发送器 22 与远程终端 11 进行通信。由计算机 31 经过发送器 22 发送的信息，被远程终端 11 中的接收器 16 接收。由远程终端 11 中发送器 15 发送的信息经接收器 21 被计算机 31 接收。

在图 1 所示的实施例中，远程终端 11 由显示器 12 、一个或几个按钮 13 、存贮器 14 、发送器 15 和接收器 16 组成。例如，装置 11 可以是三菱无线通信有限公司 MobileAccess™ 电话的类似无线电话，或是三星电子公司的 Duette 电话。在典型的无线电话中，显示器 12 是块液晶显示(LCD)屏。按钮 13 代表类似开关、键或按钮的一个或更多数据输入装置。存贮器 14 表示存贮电路或其它能存贮数字信息的装置。最好是，至少存贮器 14 的一部分是永久存贮器，意味着装置 11 关闭时信息仍被保存。在有些实施例，部分存贮器 14 组成一个一体化的推/拉高速缓冲存贮器。也可以将部分存贮器 14 作为永久存贮器或 ROM 存贮程序指令，并且装置 11 包含一个微处理器或其它类型的能执行程序指令的处理电路。

图中所示计算机 31，服务器 51 、 52，接收器 21 和发送器 22 之间的通信路径性质对本发明的实施并不重要，例如，可以利用专用和/或公共设备的交换和/或非交换的路径实现。同样，网络 40 的拓扑结构并不重要，可以由包含分级或同级网络的一系列方式实现。计算机 31 和服务器 51 可以相对于彼此安排在本地，也可以在同一个硬件上实现。

计算机 31 和装置 11 间的通信路径性质对本发明执行也不重要；然而，在很多应用上，装置 11 是如使用频率在射频至红外光谱间的电磁传输等通信技术的无线装置。应用中装置 11 是无线电话，如蜂窝电话，发送器 15 、接收器 16 、接收器 21 和发送器 22 表示用作普通电话呼叫的通信设备。

远程终端

应用时远程终端 11 和计算机 31 执行作为 HTTP 客户机功能的客户机 1，远程终端 11 至少提供三种基本功能：(1) 导航功能，允许用户导航或遍历 HTTP 统一资源定位器(URL)超链接；(2) 通信功能，与计算机 31 交换信息；及(3)界面功能，为用户提供一个向用户发送和从用户接收信息的用户界面。

最好是，这些功能可以通过采用事件驱动体系的软件控制过程实现。例如，事件可以由用户通过按钮 13 或从接收器 16 接收的信号来起始。导航功

能在两种状态中任一种状态下进行。在“准备”状态，装置等待指定超链接以遍历的用户输入；在“悬挂(pending)”状态，通信功能向计算机 31 发一个请求，则装置等待从计算机 31 接收一个回应，依据 HTTP 协议，准备状态等待被显示或处理的指定超媒体实体的 URL 的用户输入，而悬挂状态等待计算机 31 提供一个被请求的超媒体实体。

在一个实施例中，超媒体信息依据手持装置传输协议(HOTP)与计算机 31 进行交换，此协议的一个版本在由加利弗尼亚红木海岸的无线行星公司(Unwired Planet, Inc., Redwood Shores, California)于 1997 年 7 月 15 日出版、分册号为 HDTP-SPEC-DOC-101 的“ HDTP 规范”中有描述,在此列出供进一步参考。HDTP 与 HTTP 类似，但更适用于类似无线电话的远程装置，而且最好是利用用户数据电报协议/IP(UDP/IP)来传递。 UDP/IP 常常被认为没有 TCP/IP 可靠，例如，它不能保证能收到数据包，也不保证按发送的顺序接收数据包。然而，实施本发明对 UDI/IP 这类数据电报协议有兴趣，因为它不要求信息交换前在发送端和接收端建立某种“联系”。这就不需要在建立对话期间交换多个的数据包。

在一优选实施例中，超媒体信息依据手持装置标记语言(HDML)组织成卡片和卡片组。多个卡片组和其它信息实体类型可以编成所谓文摘的信息结构。这种标记语言的一个版本在由 Unwired Planet, Inc.于 1997 年 3 月出版的修订版 A 、分册号为 HDML-SPEC-DOC-200 的“ HDML2.0 规范”中有描述，在此处列出供进一步参考。

中继计算机

依据此处说明的实施例，计算机 31 和遥控装置 11 一起提供了常规超媒体客户机的功能。在此实施例，计算机 31 依据 HDTP 从遥控装置 11 接收信息，并将 HDTP 信息按需要翻译成相应的 HTTP 信息，且将结果送到服务器 51 。同样，计算机依据 HTTP 从服务器 51 接收信息并按需要将 HTTP 信息翻译成相应的 HDTP ，并将结果送到遥控装置 11 。依据本发明对计算机 31 和遥控装置 11 间交换的 HDTP 信息进行压缩从而减少了信息容量要求也减少了遥控装置 11 分析和解释信息所要求的进程。这种压缩和互补的展开过程是通过遥控装置 11 和计算机 31 内进行的编码和解码操作来完成的。

处理过程

图 2 示出了依据本发明生成文档元素压缩表达式的编码过程实施例。标

识元素 62 从路径 61 接收表示文档的信息并识别信息内的多个元素。元素常常有至少表示文档结构和语法特征某些方面的语法信息。

编码 64 生成许多代表至少某些文档元素至少一部分的多个代码。要求至少某些代码的信息容量低于所表示的元素信息要求的信息容量。代码不仅传递了所表示的元素类型也传递了元素语法信息是否存在的指示。最好是，至少一些语法信息是以降低信息容量要求的方式编码的。元素信息按需要沿路径 63 处理嵌套信息。嵌套信息可以以一系列包含递归处理过程的方式进行处理。

汇编 66 通过将编码 64 生成的代码和任何不用这些代码表示的元素或部分元素汇编成适合传输和贮存的形式从而生成沿路径 67 表示文档的编码信息。

本发明的另一个实施例包括提供许多代码簿的代码簿 68 。编码 64 自适应地从许多代码簿中选择一个代码簿并根据所选中的代码簿生成一个或多个代码。被选中的代码簿指示包含在编码信息中。

图 3 示出依据本发明从某编码表达式中复原文档元素的解码过程的实施例。标识代码 72 从路径 71 接收表示文档的编码信息并识别许多代码，每个代码至少代表相应文档元素的一部分。

根据代码，解码 74 得到语法指示并生成解码表达式。至少一些解码表达式比相应代码需要利用更多的信息容量。语法指示指示表示文档的一个或多个语法特征的语法信息是否存在。解码表达式按需要沿路径 73 去处理任何嵌套代码。嵌套代码可以用一系列包括递归过程的方式进行处理。

汇编 76 通过将由解码 74 生成的解码表达式和任何不用这些代码表示的元素或部分元素进行汇编从而生成沿路径 77 表示文档的输出信息。

本发明的另一个实施例包括提供许多代码簿的代码簿 78 。解码 74 自适应地根据编码信息内选中的代码簿指示从许多代码簿中选择一代码簿并根据选中的代码簿生成一个或更多解码表达式。

在本发明的另一个实施例中，处理 80 从路径 77 接收输出信息并沿路径 81 生成用于显示的显示信号。在某些情况下，解码 74 可能遇到不能解码的代码，因为这些代码不被解码处理认知或支持。解码 74 可以将这些不被支持的代码沿路径 73 传递给能用到这些代码的后继过程。处理 80 利用不被支持的代码中的语法指示来跳过或避免处理这些代码。

在另外的实施例中，解码 74 包含与处理 80 相类似的生成用于显示的显示信号的处理过程。在这种实施例中，因为例如显示装置不能对代码表示的元素作出适当的响应，解码 74 中的过程就采用代码所传递的元素类型和语法指示来决定哪些代码应该跳过。

编码

状态进程

编码 64 的编码过程可以利用图 4 中的状态进程进行说明，每个圆圈代表一个状态，状态间的转换用线段表示并朝箭头所指示方向转换。

编码过程起始于状态 100（开始）并沿 110 向状态 101(编码标签)转换。状态 101 生成一个相应元素标签的编码表达式。如果相关的元素标签没有伴随任何相关的语法信息，则转换沿路径 111 向状态 101 进行去生成下一个元素标签的编码表达式。如果出现一个或更多的标签属性，则转换沿路径 112 向状态 102(编码属性名称)进行。如果不存在标签属性但存在标签内容，则转换沿路 118 向状态 105(编码内容)进行。若不再出现元素标签，则转换沿路径 122 向状态 107(结束)进行从而结束编码过程。

状态 102 生成一个相关属性名称的编码表达式，有一个沿路径 113 向状态 103(编码属性值)进行的转换。状态 103 生成相应属性值的一个编码表达式。如果存在下一个标签属性，则转换沿路径 114 向状态 102 进行去生成下一个标签属性的编码表达式。

若不再有标签属性存在，则转换沿路径 115 向状态 104(属性结束)进行。状态 104 生成一个标志相关元素标签属性结束的代码。如果存在标签内容，则转换沿路径 117 向状态 105 进行从而处理标签内容。如果不存在标签内容，则转换沿路径 116 向状态 101 进行去处理下一个元素标签。

状态 105 生成一个相应标签内容的编码表达式。如果存在下一个标签内容，则转换沿路径 119 向状态 105 进行去处理下一个标签内容。若不再存在标签内容，则转换沿路径 120 向状态 106(内容结束)进行。状态 106 生成一个标志相应元素标签内容结束的代码，然后有一个沿路径 121 向状态 101 的转换从而去处理下一个元素标签。

如以下更详细地解释，标签内容可以包含嵌套元素。如果存在嵌套元素，则沿着一个未画出的路径向状态 100 进行递归转换。当嵌套的某特定层的所有元素都被处理以后，则沿着另一条没有画出的路径向状态 105 进行递归返

回转换。

例子

图 7 示出一个用如 HTML 的标记语言表达的简单文档。文档写成行的形式而且为说明方便每一行都标上序号。行序号不构成标记语言的一部分。可以预料，在实际实施例，文档传递时除了标记语言提供的以外没有任何其它行或段落的指示被传递。

行 1 包含标志一个 HTML 文档起始的<HTML>标签和行 8 包含一个标志文档结束的</HMTL>标签。在此例中，<HTML>标签没有属性但有内容，内容是分别由行 2 和行 7 上起始和结束 BODY 标签标志的文档主体。<BODY>标签的内容嵌套在<HTML>标签的内容里。表示在行 3 和行 6 之间的 BODY 标签内容包含文本和几个标签。

行 3 上的<BODY>标签内容部分表示简单的文本。行 4 上的内容部分表示这样的元素，它带有虽无内容但有指定显示图像来源的属性名称为 src 的属性及属性值(“/item.gif”)的 IMG 标签。由行 5 的内容部分表示的文本，包含一对具有用起始和结束 B 标签标明的由粗体字显示的字元素。任一个<B>标签都没有属性但每个都有文本内容。行 6 的内容部分是包含带有起始和结束A标签元素的文本，<A>标签既有属性也有内容。标签属性有名称(href)及指定其它文档 URL 的值(“ http:// a.url/info ”)。<A>标签的内容是刚好出现在结束</A>标签之前的文本“ here ”。

图 8 是将上述说明的编码过程应用在图 7 中所示的文档标记语言而得到的编码表达式的示意说明，为了便于说明，图 8 中的编码表达式排成行并标上序号，而且为了易于理解排成缩进形式。可以预料，在实际实施例中，生成的编码信息除了由标记语言元素的编码表达式提供的以外，不包含任何其它行或段的指示。

参照图 8，表示符｛XYZ-C｝表示包含标记语言标签<XYZ>的编码表达式并包含存在一个或更多标签属性及存在标签内容的指示的代码。例如，行 1 表示符｛HTML-C}表示包含<HTML>标签的编码表达式并有标签属性不存在而标签内容存在的指示的代码。同样，在行 4.1 上表示符｛IMG-A｝表示包含<IMG>标签编码表达式并包含存在一个或更多标签属性而不存在标签内容的指示的代码。

依据图 8 中所示的例子，行 1 上的表示符｛HTML-C｝表示图 7 中行 1

上表示<HTML>标签的代码。如上解释，代码传递了所表示的元素类型并包含标签内容存在的指示。行 2 上的表示符｛ BODY-C ｝表示代表<BODY>标签<图 7 中第 2 行>并指出存在标签内容的代码。

行 3 上的表示符｛ STR ｝表示用来标志文本存在的特殊代码。表示符“ The item ”代表文本本身。这个代码总暗示了存在标签内容。文本可用以显式或隐式代码的一系列方法标志。例如，文本串的开头可以通过保留文本字符中的某几个值来隐式标志。这种方案常常与上下文有关，因为这些保留值常常出现在例如二进制数据字段中。在优选实施例，文本串的开头由象图中通过表示符｛ STR ｝表示的显式代码来标志。文本串的结尾可以用象空字符或二进制零这样特殊字符来显式标志，也可以通过包含在开头代码中表示的长度值显式标志，或通过一个非有效文本字符的代码来隐式标志。对于实施本发明而言，并没有对此类特定方案的限制。

行 4.1 至行 4.3 共同表示图 7 中第 4 行上的文档元素编码表达式。行 4.1，表示符｛ IMG － A ｝表示<IMG>标签的代码并指出存在一个或更多的标签属性。行 4.2，表示符｛ src ｝表示标签属性名称“ src ”的代码。这个代码可以是下面更充分说明的其名字本身的压缩表达式或者是指明属性名称是以其它一些如传统文本串形式定义属性名称的一类属性代码。表示符(“/item.gif ”)表示给出属性值的通用文本串。或者，属性值可以编成象二进制代码的其它形式。在行 4.3，表示符｛ END:img-a ｝表示<IMG>标签的标签属性结束的代码。在本发明的一个实施例中，一个代码用来标志属性的结束，另一个代码用来标志内容的结束。在另一个实施例中，根据元素类型使用不同的代码。在另一实施例中，根据元素类型使用不同的代码标志属性和内容的结束。参照行 4.3 所示的例子，根据这些实施例，可以把表示符“ img － a ”理解成表示 IMG 属性结束标志的唯一｛ END ｝代码。然而在优选实施例中，一个象 null （空）或 0 值这样的特殊代码用来标志所有类型标签的属性和内容的结束，对于本实施例，可以把“ img － a ”理解为仅为阅读者方便而显示的标签属性、内容代码和结束代码之间的对应关系。

行 5.1 至行 5.5 共同表达图 7 中第 5 行上的文档内容的编码表达式。在行 5.1 和行 5.3 上，表示符{STR}和相伴的文本表示图 7 中第 5 行上表示两个文本串的代码和文本。

行 5.2.1 至行 5.2.3 共同表示图 7 中第 5 行上第一个<B>元素的编码表达

式。在行 5.2.1 上，表示符｛B - C｝表示<B>标签并指出存在内容的代码。在行 5.2.2 上，文本内容通过如上述的表示符｛STR｝“red”来表示。在行 5.2.3 上，表示符｛END：b - c｝表示标志<B>标签内容结束的代码。同样，行 5.4.1 至 5.4.3 共同表示了图 7 中第 5 行上第二个<B>元素的编码表达式。

在行 5.5 上，表示符｛STR｝“for a limited time”表示上述的文本串的编码并完成了图 7 中第 5 行上文档内容的编码表达式。根据图 8 中的例子，图 7 中第 6 行上的文档内容编码表达式通过行 6.1.1 至行 6.3 来共同表示。然而，在本发明的实际实施例中，相邻的文本串“for a limited time”和“Click”可以表示为{STR}“for a limited time. Click”的一个编码表达式。

正如刚才所说明的，行 6.1.1 至行 6.3 共同表示了图 7 中第 6 行上文档内容的编码表达式。如说明的，行 6.1.1 上的表示符{STR}“Click”表示文本串的编码。行 6.1.2 上，表示符{A - AC}表示<A>标签并指示存在标签属性和内容的代码。行 6.1.3 上，表示符{href}(“http:// a.url/info”)表示代表标签属性名称和值的编码。行 6.1.4 上，表示符{END:a - a}表示标志<A>标签属性结束的代码。行 6.2.1 上，表示符{STR}“here”表示标签内容的文本串的编码。行 6.2.2 上，表示符{END:a - c}标志<A>标签内容的结束。行 6.3 上的表示符表示一个文本串，从而完成了图 7 中第 6 行文档内容的编码表达式。

行 7 上的表示符{END:body - c}和行 8 上的{END:html - c}分别表示<BODY>内容和<HTML>标签结束的代码。

压缩

有各种编码或压缩方案可以用来生成信息容量要求低于用其表示的文档元素或部分文档元素的信息容量要求的代码。生成的代码既传递了所表示文档元素的类型也指出文档元素中是否存在语法信息。语法信息指示在与代码开头相应的预定的位置上传递。

根据本发明优选实施例，代码的每个字节(8 个二时制位)有固定的长度，其中一位或更多位，比如说两个最高有效位用来指示语法信息是否存在。例如，对于 HTML，两个比特分别用来指示一个或更多的标签属性和标签内容是否存在。其它代码结构可以包含可变长度的代码。例如，代码可以包含由 Huffmann 编码生成的可变长度的元素类型指示及独立的语法信息指示。语法

信息指示可以放在与代码开头相应的预定位置上。

可以根据元素类型来建立允许预定位置变化的规则，例如，该位置可以设定在紧接着一个可变长度的元素类型指示之后。在另一例子，可以预定某位置用于一类特殊代码，比如那些具有一个或几个特定值的代码，另一个位置可被预定用于其它代码。固定预定位置不依赖于在优选实施例中的元素类型。

特殊代码

在优选实施例中，建立一类六个特殊代码，这些特殊代码称为“全球代码”，因为根据这个实施例，所有编码器和解码器必须能够正确解释和处理这些代码。下面说明这六个代码。

表示符为{CBK}的特殊代码标志一个指定从众多代码簿中自适应选中的代码簿的值。根据所选中的代码簿进行解码。如上面简单解释的那样，固定长度的8位代码用来传递元素类型和语法信息指示。如果两位用来传递语法信息指示，只剩下六位传递元素类型。一般情况，元素数目远远超过6比特所能表示的数目。由于也需要用这些代码来表示经常用到的属性名称和/或属性值，这个局限就更加突出。通过将代码组成许多代码簿并从中选择合适的代码簿，编码空间的大小可以显著扩大。当一个编码器选择了一个代码簿，选择的指示就编入编码信息中因而一个互补的解码器可以决定解码应采用哪一个代码簿。{CBK}代码就是这样的指示。

表示符为{CHR}的特殊代码标志了给定字符的值。例如，用符合美国信息交换标准代码(ASCII)文本表示的文档不能表示一些定义在单一代码文本中的字府。任何单一代码字符可通过用{CHR}代码标志的数值表示。

表示符为{DAT}的特殊代码标志不能被解码器处理的“不透明”数据的起始。所谓不透明数据的含义是数据内部结构不需要被编码器知道。不透明数据标志和包含在编码信息内而无需修正。不透明数据的范围通过伴随{DAT}代码的长度值来传递。

表示符为{END}的特殊代码标志上述元素和语法信息的结束。

表示符为{STR}的特殊代码标志上述文本串的开始。

表示符为{UNK}的特殊代码标志未知元素类型。使用这个代码提高了现存编码器和解码器处理包含实现它们时未定义元素的文档的能力。过去的编码器可以以某种形式传递未知元素，即让最近的解码器去接收和处理新

元素。过去的解码器与过去的编码器一起工作能跳过以｛UNK｝代码标志的元素并继续处理其它已知代码。

解码

解码 74 的解码过程可以图 5 所示的状态过程加以说明，每个状态用一个圆圈表示。状态间的转换用线段表示并沿着箭头所指方向进行。

解码过程起始于状态 130(起始)并沿着 140 向状态 131(解码标签)转换。状态 131 生成一个从相关代码导出的相关元素标签的解码表达式。如果相关代码指示不存在语法信息，则转换沿路径 141 向状态 131 进行从下一个代码生成解码表达式。如果代码指示存在一个或更多的标签属性，转换沿路径 142 向状态 132 进行(解码属性名称)。如果代码指示不存在标签属性但存在标签内容，转换沿路径 148 向状态 135(解码内容)进行。若不再出现元素标签，转换沿路径 152 向状态 137(结束)进行从而结束解码过程。

状态 132 生成一个相关属性名称的解码表达式。转换沿路径 143 向状态 133(解码属性值)进行。状态 133 生成相应属性值的解码表达式。如果存在下一个标签属性，转换沿 144 向状态 132 进行从而生成下一个标签属性的解码表达式。

当不再出现的标签属性时，如果存在标签内容，转换沿 147 向 135 进行去处理标签内容。如果不存在标签内容，转换沿路径 146 向状态 131 进行去处理下一个代码。

状态 135 生成相关标签内容的解码表达式。如果存在下一个标签内容，转换沿路径 149 向状态 135 进行去处理下一个内容。若不再出现另外的标签内容，转换沿路径 151 向状态 131 进行去处理下一个代码。

正如上述，标签内容可以包含嵌套代码。如果存在嵌套代码，则沿着一个没出标出的路径向状态 130 进行递归转换。当嵌套的某特定层的所有代码都被处理以后，则沿着另一条没有标出的路径向状态 135 进行递归返回转换。

递归

图 4 和图 5 示出的状态图没有显示任何递归的规定。实施本发明并不要求递归，但在处理嵌套元素和代码的许多实施例中却是一种有效的技术。图 6 中所示的功能流程图表示了编码或解码用 HTML 这样的标记语言表达的文档元素的递归过程。

编码

根据所示的编码过程，步骤 221 完成各种初始化任务。步骤 222 将循环级数初始化为零。步骤 223 处理元素标签从而生成编码表达式。步骤 224 检查是否存在任何标签属性。如果有，步骤 225 处理标签属性从而生成一个编码表达式，然后返回步骤 224 检查是否存在其它标签属性。若不再出现其它标签属性，处理过程继续到步骤 226 。

步骤 226 检查是否存在标签内容。如果有，步骤 227 处理标签内容从而生成一个编码表达式。步骤 228 检查标签内容内是否嵌套了任何元素。如果没有，处理过程返回到 226 去检查是否存在任何其它标签内容。若不再出现其它标签内容，处理过程继续至步骤 230 。如果标签内容内嵌套了一个元素，步骤 229 递增循环级数并继续执行步骤 223 。

步骤 230 检查当前循环级数是否为零。如果不为零，步骤 231 递减循环级数并继续处理步骤 226 的执行过程。如果循环级数为零，步骤 232 检查是否完成了编码过程。如果没有，处理过程返回至步骤 223 。如果完成了编码过程，步骤 233 执行各种结束任务。

解码

根据所示的解码过程，步骤 221 完成各种初始化任务。步骤 222 将循环级数初始化为零。步骤 223 处理代码从而生成一个解码表达式。步骤 224 检查是存在任何标签属性。如果存在，步骤 225 处理表示标签属性的代码从而生成一个解码表达式；然后返回至步骤 224 检查是否存在任何其它标签属性。若不再存在其它标签属性，处理过程继续至步骤 226 。

步骤 226 检查是否存在标签内容。如果存在，步骤 227 处理表示标签内容的代码从而生成一个解码表达式。步骤 228 检查是否有任何代码嵌套在经编码的标签内容内。如果没有，处理过程返回至步骤 226 检查是否存在任何其它标签内容。若不再存在其它标签内容，处理过程继续至步骤 230 。如果有一个代码嵌套在经编码的标签内容内，步骤 229 递增循环级数并继续步骤 223 的处理过程。

步骤 230 检查当前循环级数是否为零。如果不为零，步骤 231 递减循环级数且处理过程继续至步骤 226 。如果循环级数为零，步骤 232 检查是否已完成解码过程。如果没有，处理过程返回至步骤 223 。如果已完成解码过程，步骤 233 执行各种结束任务。

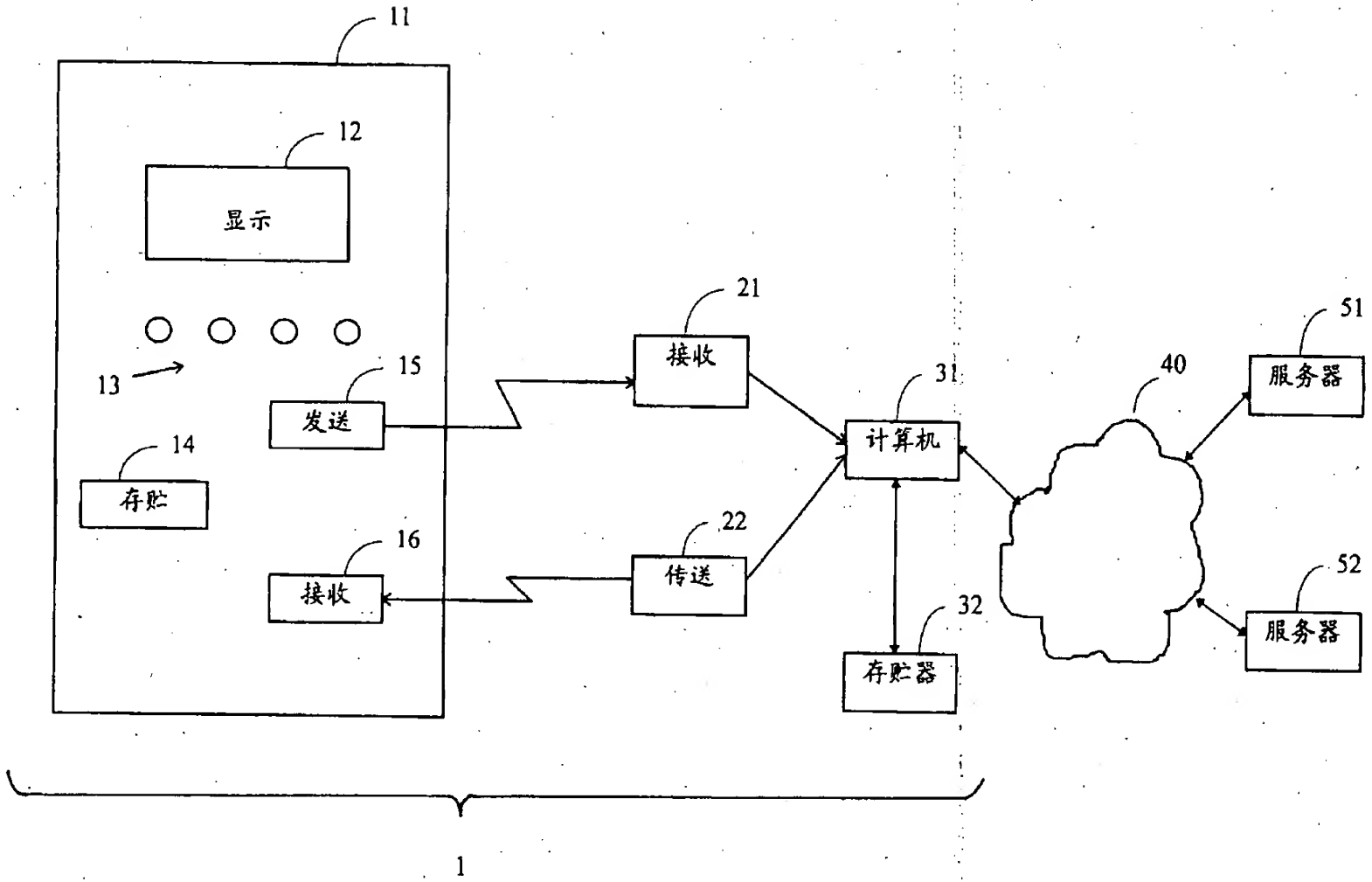

11
12
显示
13
14
存贮
15
发送
16
接收
1
21
接收
22
传送
31
计算机
32
存贮器
40
51
服务器
52
服务器

图 1

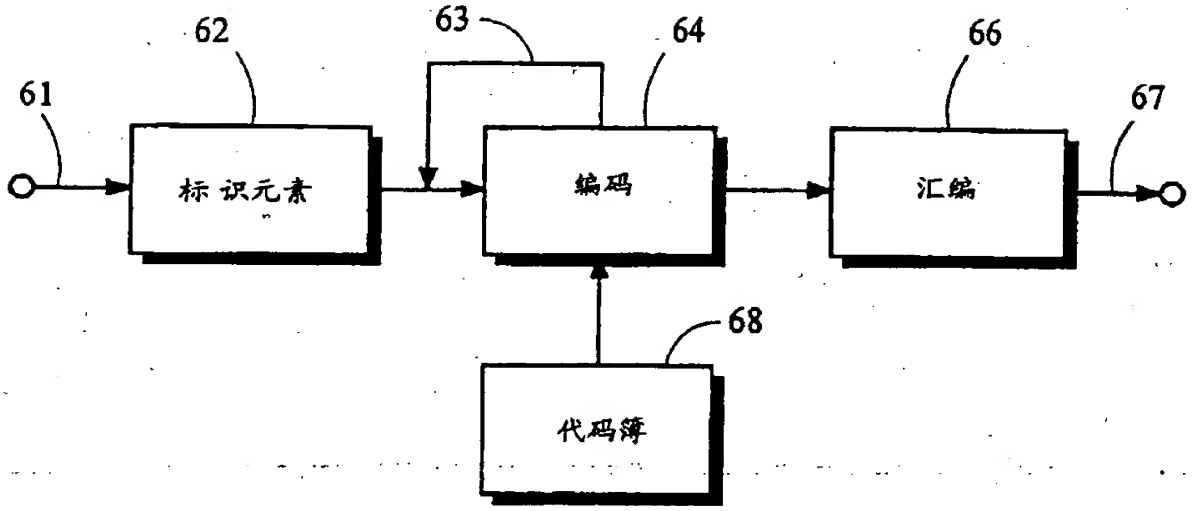

61
62
63
64
66
67
标识元素
编码
汇编
68
代码簿

图 2

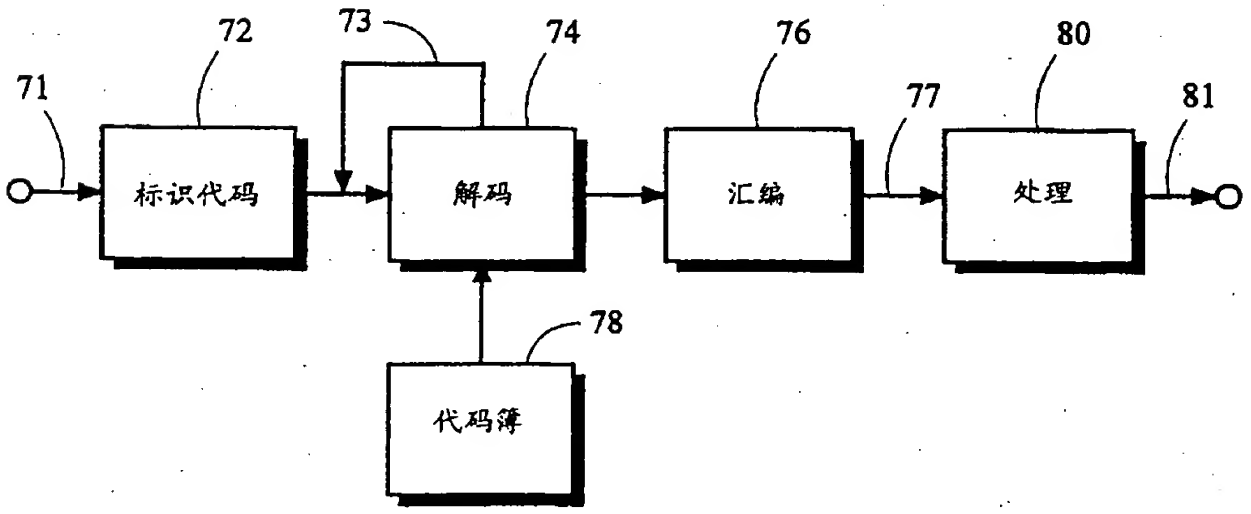

71
72
73
74
76
77
80
81
标识代码
解码
汇编
处理
78
代码簿

图 3

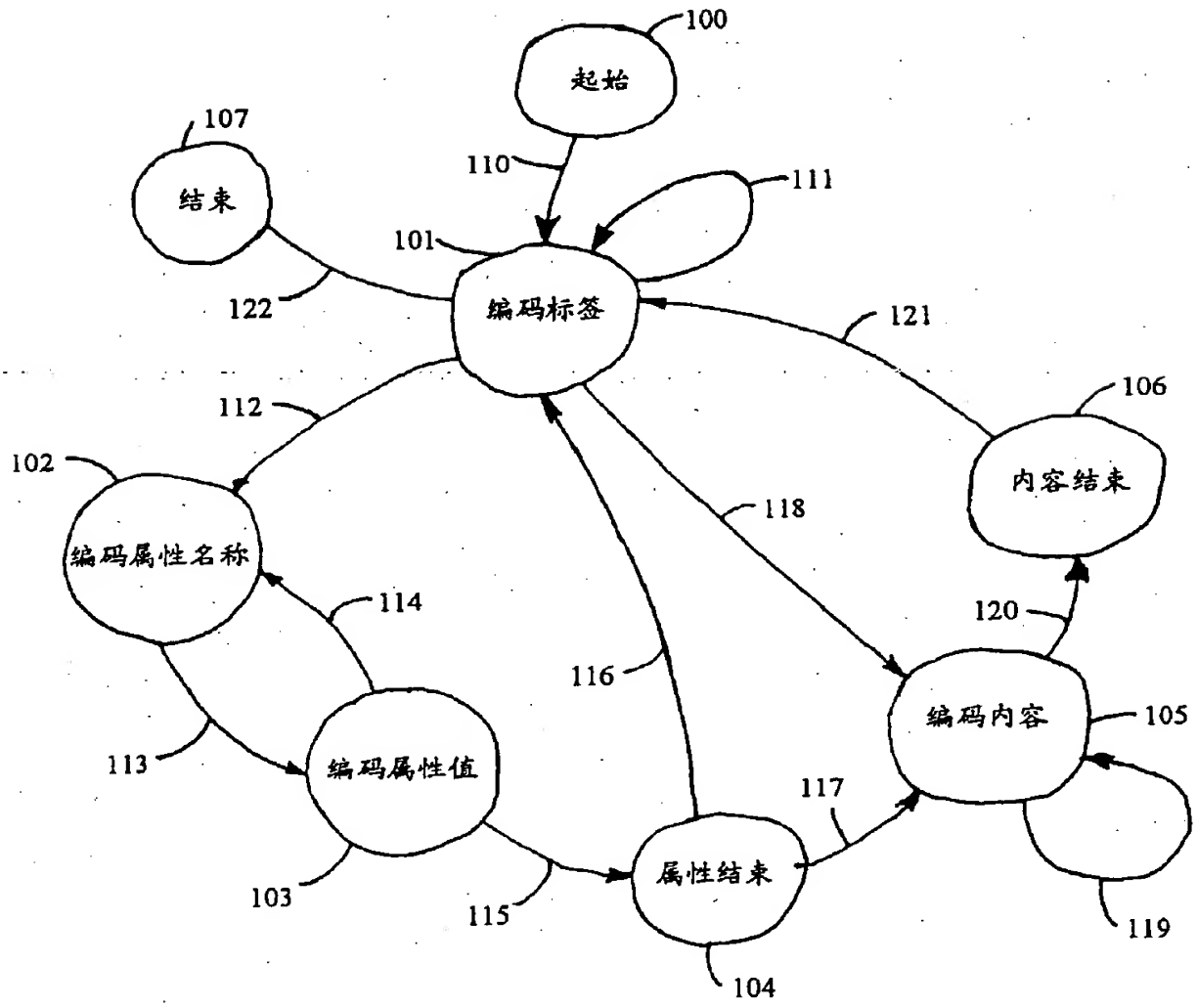

100
起始
107
结束
110
111
101
编码标签
122
121
106
112
内容结束
102
编码属性名称
118
114
120
116
编码内容
105
113
编码属性值
117
属性结束
103
115
119
104

图 4

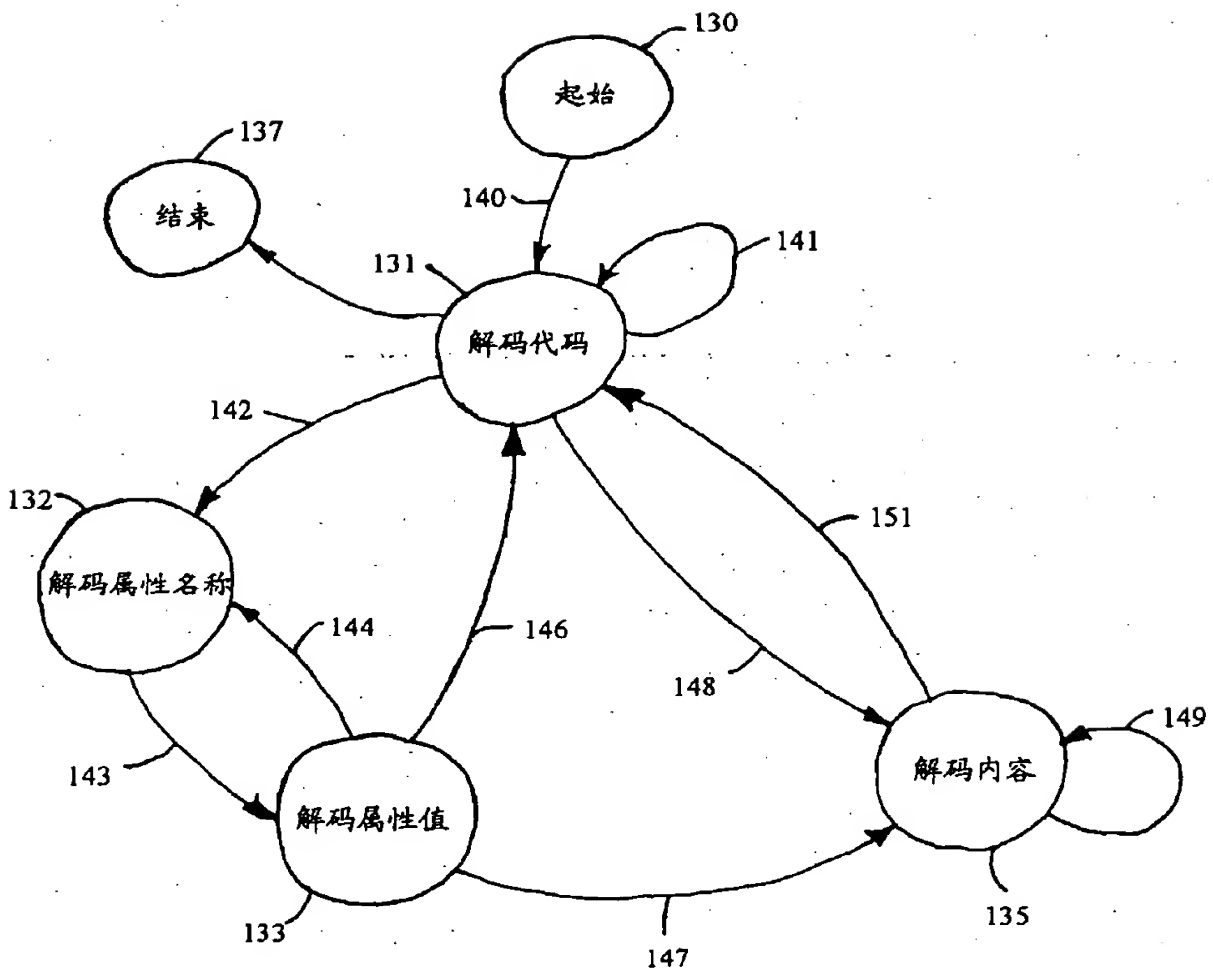

130
起始
137
结束
140
131
141
解码代码
142
132
解码属性名称
144
146
151
148
149
解码内容
143
解码属性值
133
147
135

图 5

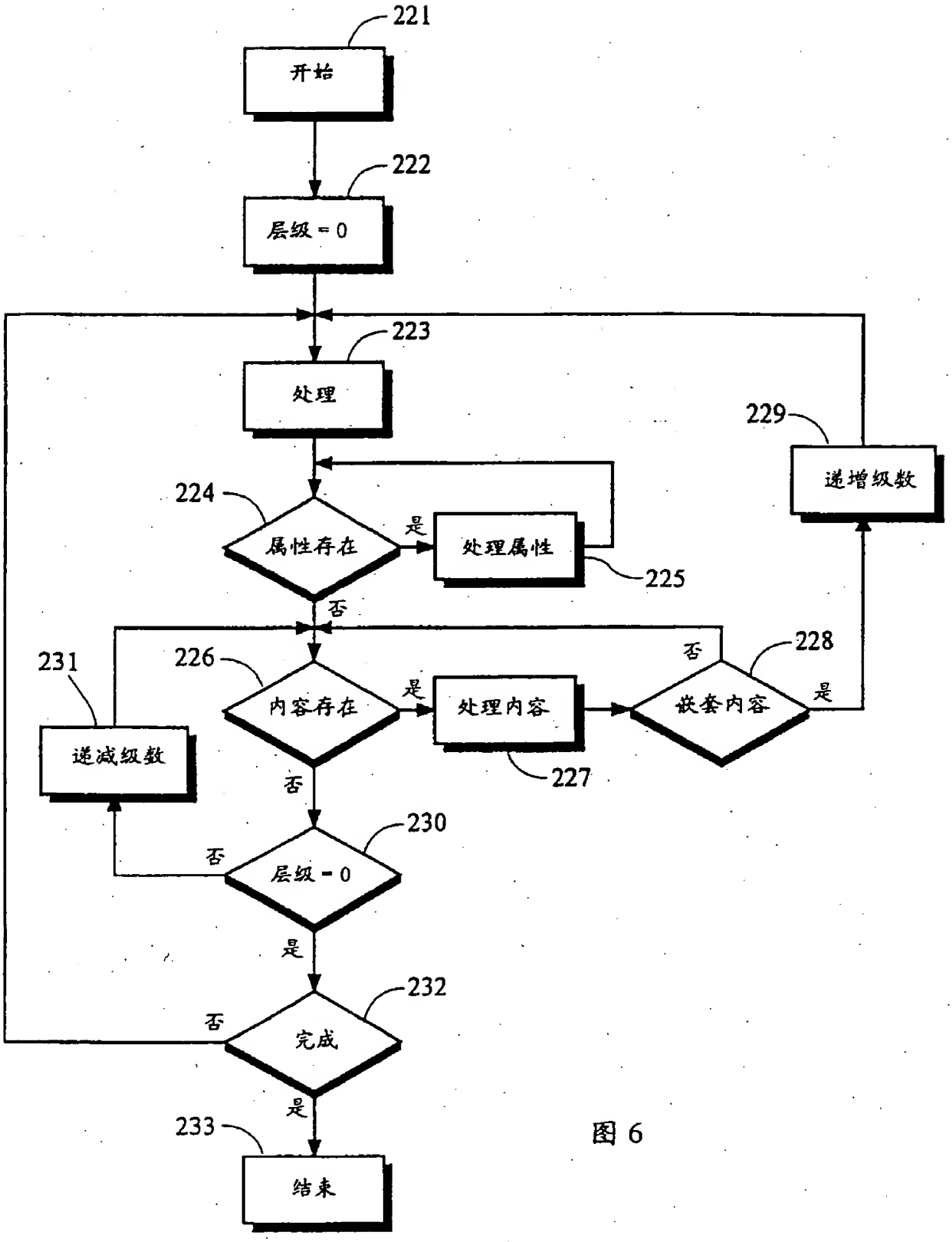

221
开始
222
层级 = 0
223
处理
229
递增级数
224
属性存在
是
处理属性
225
否
231
226
否
228
内容存在
是
处理内容
嵌套内容
是
递减级数
227
否
230
否
层级 = 0
是
232
否
完成
是
233
结束

图 6

```
1. <HTML>
2. <BODY>
3. The item
4. <IMG src="/item.gif">
5. is available in <B>red</B> and <B>green</B> for a limited time.
6. Click <A href="http://a.url/info">here</A> for more information.
7. </BODY>
8. </HTML>
```

图 7

```
1.    {HTML-C}
2.        {BODY-C}
3.            {STR} "The item "
4.1           {IMG-A}
4.2               {src} ("/item.gif")
4.3               {END: img-a}
5.1           {STR} "is available in "
5.2.1         {B-C}
5.2.2             {STR} "red"
5.2.3             {END: b-c}
5.3           {STR} " and "
5.4.1         {B-C}
5.4.2             {STR} "green"
5.4.3             {END: b-c}
5.5           {STR} " for a limited time.  "
6.1.1         {STR} "Click "
6.1.2         {A-AC}
6.1.3             {href} ("http://a.url/info")
6.1.4             {END: a-a}
6.2.1             {STR} "here"
6.2.2             {END: a-c}
6.3           {STR} " for more information."
7.            {END: body-c}
8.        {END: html-c}
```

图 8